**Projet de cartographie du Web littéraire francophone**

# Présentations des étudiants

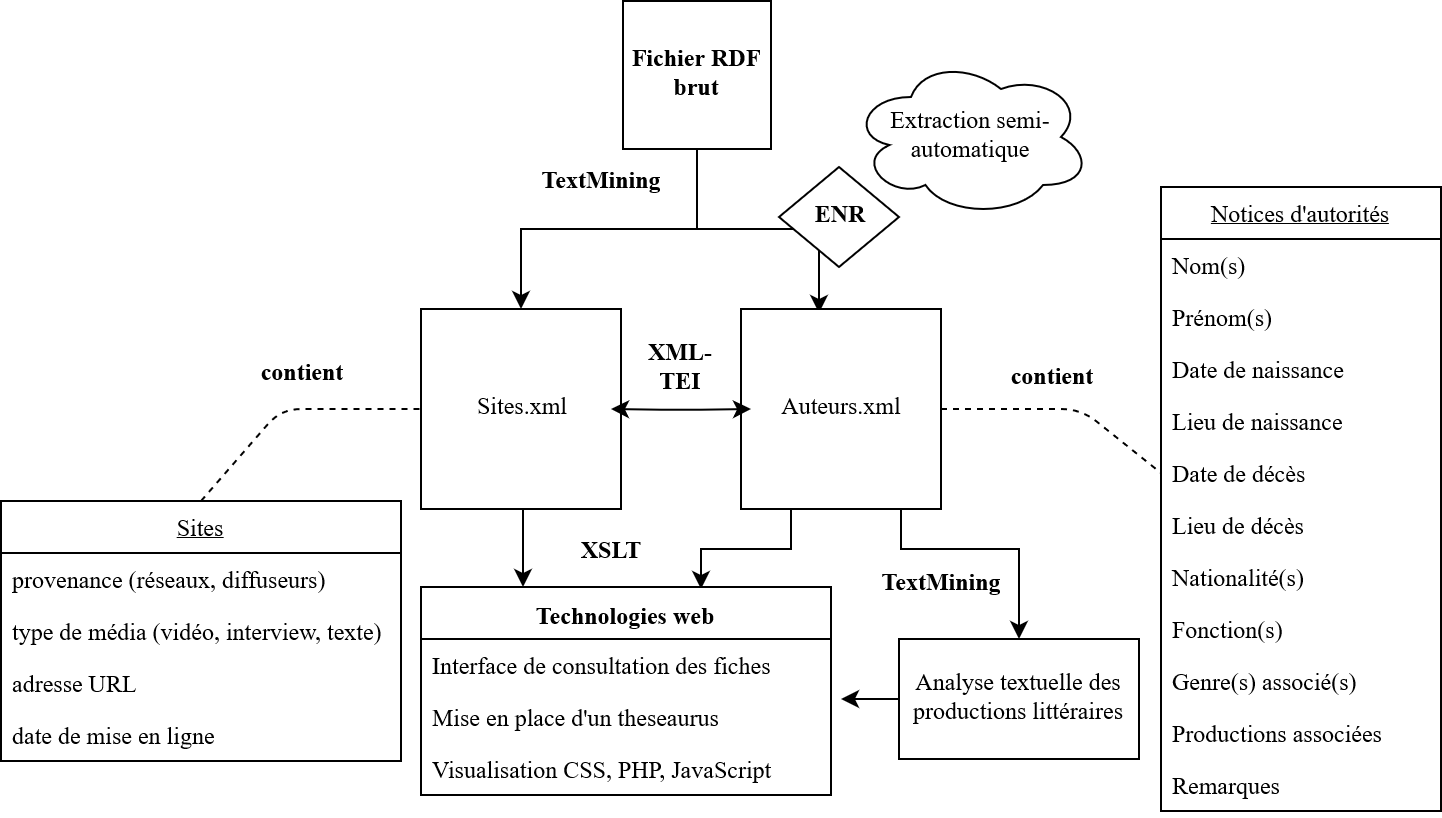
**Inès Burri :** Diplômée d’une licence de Lettres Classiques à l’université Jean Moulin Lyon 3, et actuellement en Master 2 Mondes anciens — Histoire ancienne. J’étais, avant ma reprise d’études, infographiste 2D-3D et je mets en pratique ces connaissances pour mon mémoire de recherche sur la localisation de la bataille de *Lugdunum* en 197 ap. J.-C. J’ai également effectué un stage l’année dernière dans le projet Hyperdonat, une édition numérique d’une traduction française, d’annotation et de critiques des commentaires de Donat aux comédies de Térence à l’aide de l’indexation XML-TEI.

**Anaïs Chambat** : Dans le cadre de mon master de recherche en histoire contemporaine des mondes germaniques (2017, Strasbourg), j’ai travaillé sur les modalités de diffusion des savoirs statistiques et géographiques allemands. Je poursuis actuellement ma formation à l’ENS de Lyon où je m’intéresse aux applications du numérique et des méthodes quantitatives à l’histoire. Mes orientations de recherche portent notamment sur les *mediums* de la propagande économique du XIXe siècle. Je mobilise pour cela divers outils heuristiques tels que la prosopographie et les méthodes de l’histoire quantitative (analyse textuelle conjointe sous IRaMuteQ et TXM, annotations XML).

**Célian Ringwald :** Diplômé d’un master en science des données à Lyon 2, je travaille en tant que *Data scientiste* depuis 3 ans à Marseille chez *Data Observer*, une entreprise spécialisée en vieille d’informations médiatiques. J’y développe des solutions logicielles, visuelles et algorithmes permettant l’analyse, l’annotation et la consultation de corpus dynamiques ciblés sur des problématiques informationnelles d’entreprises ou d’institutions. Afin de pouvoir réaliser un tournant académique, j’ai repris une formation en humanités numériques à Lyon que je suis pour l’instant en parallèle de mon activité professionnelle à Marseille.

# Pistes de travail

Schéma conceptuel des pistes de travail réalisé sous Draw.io



1. Procéder à un état de l’art sur les outils de visualisation et sur les « données liées »
2. Enrichissement des métadonnées existantes
   * Création de notices d’autorités
   * Mise en place d’un *thesaurus* par mots-clés
3. Mise en place d’une base de données pour verser automatiquement les données
   * Prototypage d’annotations XML-TEI sur un corpus restreint de productions littéraires
   * Analyse textuelle (fouille de texte, IRaMuteQ, TXM, Hyberbase)
4. Visualisations visant à proposer des parcours de lecture originaux à partir de l’indexation ou de l’analyse établie en amont
   * Mobilisation de technologies web de type HTML, PHP, CSS et JavaScript
   * Analyses de réseaux (Gephi)
     + Pistes d’explorations et de visualisation de contenu par auteurs, pays, période, types de médias, genres littéraires et thématiques d’étude, etc.

# Reformulation de la commande

Attente n° 1 : compréhension du matériau mis à disposition par le projet de recherche (plusieurs centaines de sites sont déjà disponibles)

À la suite de notre rendez-vous avec Julien Velcin le lundi 23 septembre 2019, nous avons pu discuter du format et du contenu des données que nous aurons à traiter. Nous avons ainsi pu prendre connaissance du fichier RDF et de la suite des liens.

Nos questions :

* Y a-t-il une seule production littéraire par page web ?
* Est-ce un échantillon aléatoire ou est-il fermé sur une thématique « définie » ?
  + A-t-on un registre d’auteurs prédéfinis auquel se référer ?

Attente n° 2 : mise en place d’une base de données et des outils pour verser automatiquement les données issues d’un nombre limité de pages Web au format clairement défini

Nos questions :

* Doit-on réaliser une interface de consultation ?
* D’import de documents manuel ou « automatique » ?
* Doit-on automatiser la création d’une fiche (avec les métadonnées basiques) et de son classement (classement par auteur, par genre littéraire, par date, par pays) ?
  + Si oui, y a-t-il un format défini (rdf, xml, xslt) ?

Attente n° 3 : réflexion sur les visualisations les plus adaptées à un utilisateur non expert

Nous avons actuellement quatre entités à représenter : site web — date — auteur — type de média utilisé. Pouvons-nous intégrer d’autres informations déductibles des documents ?

Attente n° 4 : implémentation d’une ou plusieurs de ces visualisations (si possible)

Nos questions :

* Quels sont les besoins identifiés ?
  + Une visualisation dynamique ?
    - Analyses de réseaux ;
    - Cartes thématiques ;
  + Cartes de localisation (type SIG) ?
  + Pour quelle utilisation ?
    - Découverte ;
    - Analyse ;
  + Ces analyses doivent-elles être *offline* ou *inline*
* Y a-t-il des *mediums* privilégiés pour la vulgarisation des données (site Web, poster) ?